Applications of Omnidirectional Imaging:
Multi-body tracking and remote reality

Terry Boult tboult@eecs.lehigh.edu
Chen Qian, Weihong Yin, Ali Erkin, Peter Lewis, Chris Power, Ross Micheals
Vision and Software Technology Lab, EECS Department Lehigh University

Abstract

Recently, S. Nayar introduced a parabolic imaging system that has a field of view of a full hemisphere or more. When used with a video camera the result is an omni-directional video stream that captures everything going around it. In the VAST Lab at Lehigh, we have been experimenting with these cameras, developing new variants, and developing omni-directional vision applications.

We present an overview of omni-directional imaging and then two of our applications which we will be demonstrating. The first application is a frame-rate multi-body tracking system. The system uses an omni-directional imager and a standard PC to track multiple moving objects in all directions. The system is designed to provide perspective views of the most significant targets, either locally or over a network.

The second application is something we call Remote Reality, which provides an immersive environment via omni-directional imaging. It can use live or pre-recorded video from a remote location. While less interactive than traditional VR, remote reality has important advantages: there is no need for "model building" and the objects, textures and motions are not graphical approximations.

1 Omni-Directional Imaging

Recent research, [**?**], has revolutionized wide-field of view imaging by introducing the paracamera, a system that directly captures a full hemisphere (or more) while maintaining a single perspective viewpoint. Because it captures the viewing hemisphere (or more) simultaneously it be used for full motion video. Furthermore, placing two paracamera systems back-to-back allows a true viewing sphere, i.e. 360 x 360 viewing. Unlike fish-eye lenses, each image in the paracamera system can be processed to generate *geometrically correct* perspective images in any direction within the viewing hemisphere.

The omni-directional imager combines an orthographic lens and a parabolic mirror, where the axis of the parabolic mirror is parallel to the optic axis. Because the lens is orthographic, entering rays are parallel. By definition, rays parallel to the axis reflect off a parabolic surface at an angle such that they virtually intersect at the focus of the parabolic surface. Thus the focus of the paracamera provides a single "virtual" viewpoint. The single virtual viewpoint allows for consistent interpretation of the world in any viewing direction. To generate a proper perspective image from the para-image, consider an "imaging array" in the desired viewing direction. For each pixel, it logically casts rays through the focus and intersects the measured image. The resulting spatially varying resampling can be very efficiently implemented using spatial lookup tables.

The parabolic mirror and orthographic imaging system was patented by S. Nayar and a commercial version, called the paracamera is available from www.cyclovision.com (with basic WindowsNT software). The stock paracamera is self-contained. Smaller, custom designs for camcorders can be produced. We have built a few of our own for use in our research projects, e.g. see figure 1.

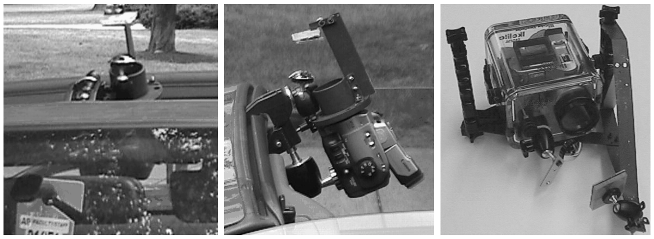

Figure 1: Some custom omni-directional cameras for vehicle use and underwater use.

Because omni-directional imaging compresses a viewing hemisphere into a small image, maintaining resolution and captured image quality is quite important. While the process scales to any size imager, the current systems use NTSC (640x480) or PAL (756x568) cameras. The spatial resolution along the horizon is $\frac{240\text{pixels}*2*\pi}{360\text{degrees}} = 4.2\frac{\text{pixels}}{\text{degrees}}$ (5.1 for PAL). Note the "spatial resolution" of the image is not uniform. While it may seem counter intuitive, the spatial resolution of the omni-directional images is *greatest* along the horizon, just where objects are most distant.

2 The Remote Reality System

The main components of the remote reality system are the omni-directional camera, video recording systems, car mounting bracket and a head-mounted-display (HMD). Remote Reality , is a very simple application. The head tracker in the HMD provides orientation information which determines the unwarpping map. As the HMD turns or one "zooms in" the virtual viewpoint is stationary; only the virtual "imaging array" is moved. Making the system fast took a few, but straight forward tricks: fixed point math for most computations and table lookup for expensive operations. It also uses an efficient 4-case factor-of-2 bilinear-based interpolation done directly in RGB555 color space with no multiplications.

The prototype system balances cost and quality. Our current data collection system was approximately $4K (+$1K for underwater) and the computing/HMD play-back system was about $3K. The system uses a 233MMX CPU (running Linux) & video capture card. The system computes monocular SIF-resolution full-rate "video" (320x240 30 fps NTSC)

Figure 2: Omni-directional camera on car and Remote Reality Driver.

which is reasonably well matched to the Virtual I-O glasses. The built in head tracker provides yaw,pitch and roll, with updates to the viewing direction at 15-30fps. (A mouse or joystick can also be used for view selection.) We are currently adding GPS localization to the collection system to better support augmented reality applications.

A natural application of remote reality is in remote vehicle operation. We have mounted a paracamera on a radio-controlled car and the remote driver wears the HMD as he/she drives, see figure /reff:RR. We are currently building, and hope to demonstrate, a wearable version of the system to allow the driver to follow the remote vehicle as one might in a dangerous situation.

3 Frame-rate tracking

A second natural application of an omni-directional sensor is for surveillance. By seeing a full hemisphere, there are no blind spots and no need for panning back and forth. This makes it ideal for a background subtraction based tracking. The difficult part of omni-directional tracking is the resolution, since the full hemisphere of view targets will be quite small; using NTSC video, a human (2m by 1m) at 50 meters will project to 20-30 total pixels. Therefore we cannot, as many frame-rate trackers do, reduce the resolution of the image.

Our system has 6 main components:

1. background adaption
2. background subtraction
3. connection and labeling
4. temporal association
5. significance sorting
6. target display

The background adaption is to help reduce the impact by slowly varying lighting changes. A rapid adaption, as many systems use, reduces the ability to track slowly moving targets and targets that are directly approaching the sensor. Our background adaption is slow, usually adding in 1/8 of a new image every 16-128 frames. Thus a target with an intensity difference of 64 gray levels would take 2-30 seconds to be below detection threshold. The infrequent updates not only allow detection of more slowly moving targets, it also reduces the computation demands. The background is both a mean and variance for each pixel, and we have been exploring a multiple background model as well. This is still an area of ongoing development.

The current background subtraction algorithms use MMX to speed up the actual computation. Each difference is compared to the sum of its variance value and a user-variable threshold. To handle rapid lighting changes, the system detects if too large a fraction of the pixels are above threshold and, if so, temporally increases the threshold. For speed we maintain, per row, information on the first and last pixels in the row that were above threshold.

Forming connected components and labels is the first time our system reduces resolution, which is reduced by a factor of 4 in each direction. However, to maintain the sensitivity each “parent-pixel” in the lower resolution label image maintains a count of how. Since most of the pixels are not above threshold we skip empty rows, and for non-empty rows we use integer pointers (4 bytes/pixels at a time) until we locate actual pixels that were above threshold. The connectedness is applied at the lower resolution, thus small gaps are generally filled. A second pass relabels the image and removes regions whose total areas were out of range.

The temporal association has 3 components. The first is done during the connected component labeling where the label image from the previous image is checked, per low resolution pixel, to make suggested temporal associations. Then, for each unassociated region, nearby regions in the previous frame are compared for similar size with the closest matching objects being postuated as associated. This is also a search over unassociated targets in previous frames to handle short (temporally) occlusions. This is still an area of ongoing development.

Significance sorting is still being investigated, but for now we take a weighted average of target age (in frames), size in pixels and time since last “display”. In the current GUI, boxes surround each target on the local display of the paraimage. In addition the N most significant targets can then be perspectively unwarpped and displayed in separate windows.

We are currently working on distributed scheduling and a display module where multiple omnidirectional systems interact and share a common scheduling/display system. The system is parameterized for both scheduling and display parameters and includes network bandwidth computations/limitations.

4 Conclusion

Omnidirectional imaging has opened new doors for computer vision applications. This paper presented a summary of the techniques used in two of our demonstrations: remote reality and frame-rate multi-body tracking. Both of these will be interactive demonstrations at the workshop on applications of computer vision.

[Nayar, 1997] S. K. Nayar. Catadioptric Omnidirectional Camera. *Proc. of IEEE CVPR*,June 1997.